Business Objective

The objective is to provide recommendations to TourRadar's management on which version of the homepage to use.

Data

The data are the results of an A/B test performed on the TourRadar platform in October 2017. It has 2430 observations and 9 variables. These variables include: date, version, location, device, browser, visit, bounce rate, time on page and conversions. The data dictionary is shown in table 1. There was no missing data in the provided dataset.

Table 1: Data Dictionary

Variable	Variable Definition
date	date of the observation
version	version of the page: 0 = original homepage without a search bar, 1 = homepage with a search bar only, 2 = homepage with a search bar and a banner
location	location of the visitor
device	device type of the visitor
browser	visitor's browser name
visits	the number of unique visits on this particular version of a homepage
bounce_rate	bounce rate (the percentage of single-page sessions in which there was no interaction with the page)
time_on_page	the average amount of time (in seconds) users spent viewing a specified version of a homepage
conversions	the number of bookings

Analytics Approach

The approach would be to use exploratory analysis and statistical modeling. Exploratory analysis would be used to get descriptive statistics, high level observations and correlations. Statistical analysis will be used to choose the best version.

Exploratory Analysis

The descriptive statistics is shown in table 2. It shows the date range is between October 1, 2017 to October 18, 2017. The maximum number of visits was 3235. The average bounce rate was 0.4444. Average conversion was 2.168 and average time of page was 2.5 minutes.

Table 2: Descriptive Statistics

	date	visits	bounce rate	time_on_page	conversions
Minimum	2017-10-01	7	0.32	100	0
1st Quartile	-	113	0.38	125	1
Median	2017-10-09	211	0.44	150	1
Mean	-	297.7	0.4444	150.3	2.168
3rd Quartile	-	376.5	0.51	176	3
Maximum	2017-10-18	3235	0.57	200	36

There seems to be a medium positive correlation between conversions and visits. This is shown in figure 1.
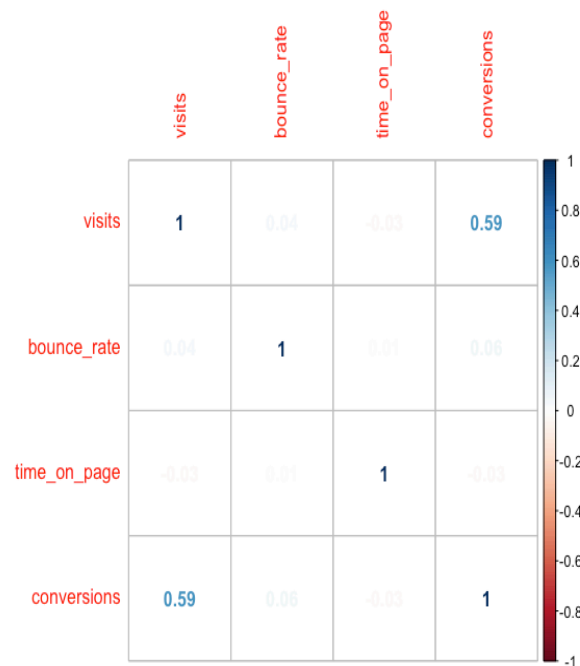

Figure 1: Correlation plot between Visits, Bounce rate, Time on page and conversions

Figure 2 shows that version 0 had a high conversion rate compared to version 1 and 2. This will be validated in the statistical modeling section.
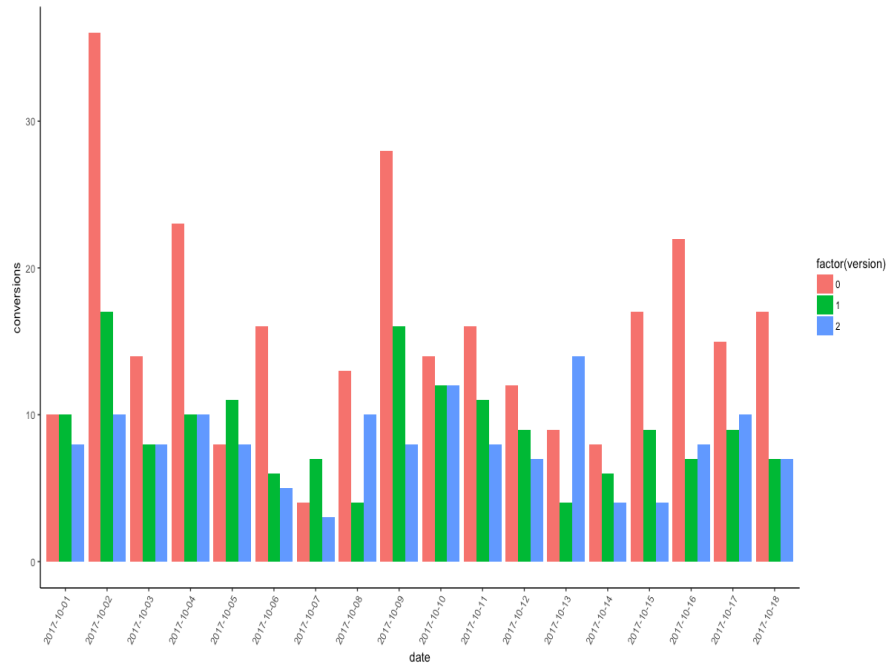
Figure 2: Bar plot of date, version and conversions

Figure 3 shows the spike in visits coincides with the conversion rate. The bounce rate and time on page seem to follow a similar pattern.
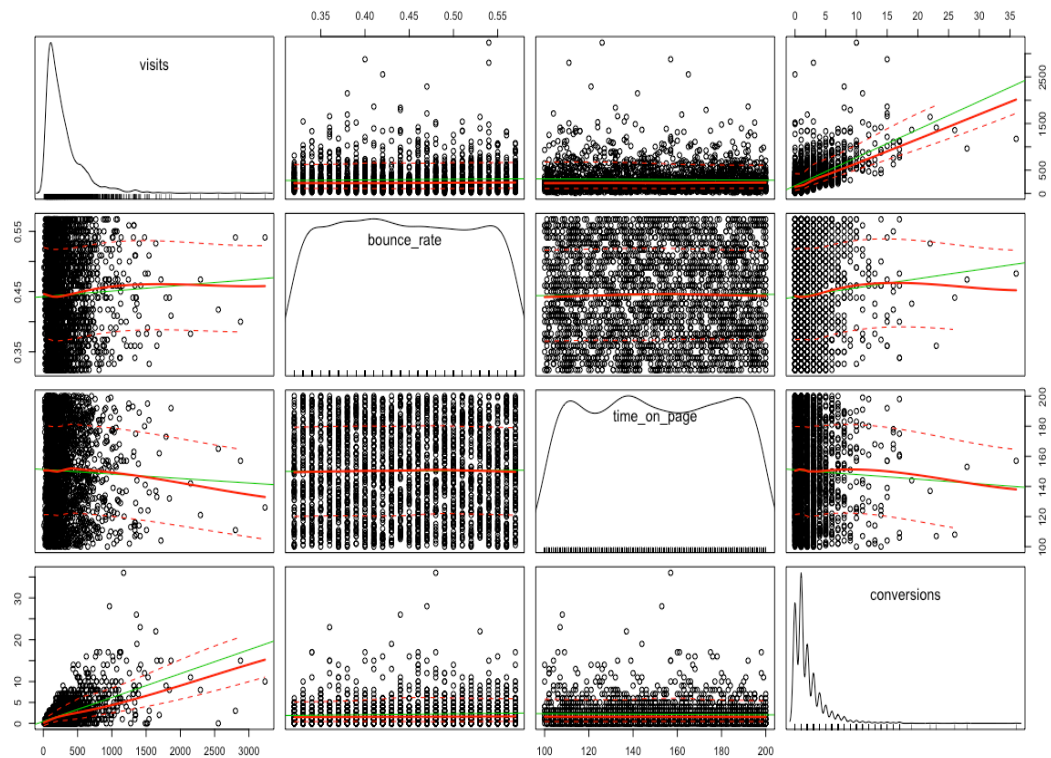

Figure 3: Scatterplot matrix for visit, bounce rate, time on page and conversions

Statistical Modeling

Let's first look at proportionality. In figure 4, it shows the comparisons between versions 1 & 2, 1 & 3 and 2 & 3. Figure 4 shows that version 1 has the highest proportion among all three versions.
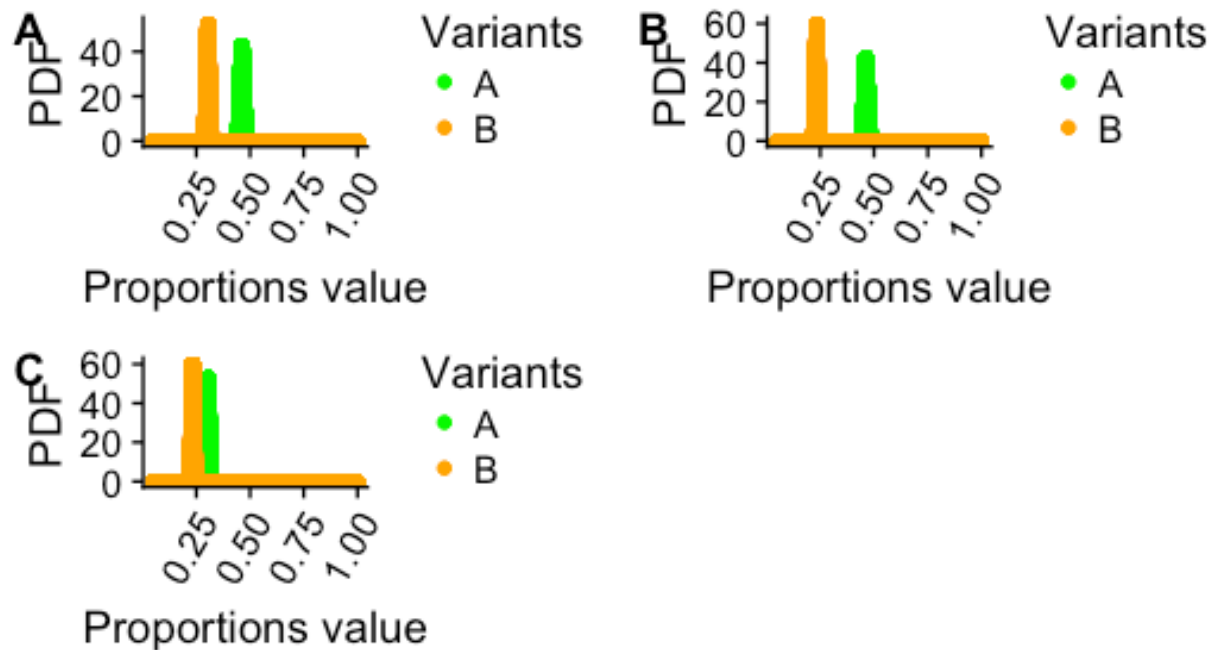

Figure 3: Proportion Comparisons between the different versions

We can then use hypothesis testing to validate which one which should choose. The hypothesis is that the true conversion of our visitors is variant A and that the proportion pb we collected in the B variant is simply due to chance. It would be done three times for the three combinations [(V1, V2), (V1, V3), (V2,V3)].

After calculating the type1 and 2 errors of both it shows that there was no significance when V2 and V3 were run.

Insights

Based on the analysis here are the key takeaways.
- Choose version 1 as having the new features has no impact on conversions
- Use more data for the process as it is a pretty small time-frame

Next steps

See if we can perform the test over a larger period of time in order to get a richer dataset. The process should be performed multiple times to verify the results. From a technique perspective applying Bayesian A/B testing may yield better results. The approach is fairly new but could be useful technique in future. Lastly, build an AB testing framework to make the process easily repeatable.